CHAPTER 8

EQUILIBRATION OF TRAVEL DEMAND AND SYSTEM PERFORMANCE: AN APPLICATION IN A TRANSPORTATION CORRIDOR[1]

Introduction

Equilibration of travel demand and system performance has traditionally been considered an important problem, primarily for auto travel, and only secondarily for bus travel. Initially, "capacity-constrained traffic assignments" were predominant. These algorithms were invariably restricted to auto mode, provided no information on convergence, and suffered from other shortcomings (Ruiter, 1973). In short, they were not valid equilibration procedures.

In the last decade or so substantial progress has been made in the development of transportation network equilibrium algorithms. The seminal paper of Dafermos and Sparrow (1969) can be identified as a turning point: it signalled the start of mathematically rigorous analyses of the equilibration problem. For a unified approach to network equilibration methods, viewed as solutions to an optimization problem, the paper by Nguyen (1974) is definitive. A brief exposition of the network algorithms is given in the second section of this chapter.

[1]The "forerunner" of this chapter, Working Paper Number 7621, by Ibrahim Hasan and Antti Talvitie, was published in Proceedings of the World Conference in Metropolitan Research, Rotterdam, 1977.

The objective of this chapter is to suggest a way in which an equilibrium model can be developed, where the major components are a set of transportation system performance equations and a set of disaggregate travel demand equations. The method, which utilizes the Scarf (1973) algorithm, will be developed within a context of an analysis of work-trip demand and mode-choice over a transportation corridor.

An experienced transportation planner readily recognizes that this is a limited application. It is the authors' belief, however, that over the next few years our approach to predicting travel demand and computing equilibrium will prove to be a useful complement to network equilibrium models. We also believe that there will be increasing use of this approach in applications of greater complexity. This chapter outlines the components of the proposed system, summarizes the key concepts involved in the algorithm, and provides an example of its application. A "real world" application and computational experiences are discussed in Talvitie, *et al.* (1977) .

Equilibration in Transport Networks

This section provides a heuristic description of network equilibration methods. The emphasis is on simplicity, not on mathematical elegance, and only the "user-optimized" equilibrium problem will be considered.[1] The presentation here is not a substitute for more complete treatments of network equilibrium (Nguyen, 1974).

Basic concepts

Before formally stating the network equilibrium problem, it is instructive to briefly review why equilibration in networks is difficult. The problem is the following. The demand functions are expressed in terms of origin-destination pairs, while the network performance functions are facility (link)-specific. Mathematically,

$D_w = d_w(C_w)$, the demand function for O-D pair w , and

$C_a = c_a(D_a)$, the link performance function on link a , where

C is "generalized" cost (travel time, cost, etc.) and D is demand.

Thus, the interzonal travel time that enters the interzonal generalized cost C_w depends on the path taken by the traveler. In networks, this O-D travel time is calculated as the summation of the individual link travel-times forming the path. On the other hand, the link travel times are functions of the link volumes, which are composed of flows between many origins and destinations. Therein lies the problem: the "markets" for demand are the O-D pairs when the "markets" for supply are links.

[1]The difference between "user-optimized" and "system-optimized" equilibria is discussed in several places. A good discussion is given by Ruiter (1973).

Most origins and destinations are also connected by many paths, and a rule or model of path choice behavior must be given for the equilibrium volumes to be unique. The user-optimized equilibrium is governed by Wardrop's (1952) First Principle, which states, "the journey times on all the routes actually used are equal, and less than those which would be experienced by a simple vehicle on any unused route."

Relationships for network equilibrium

Armed with the basic understanding of the problem, we can state it more formally. The following notation will be used:

$N = (n_1, n_2, ...)$,	set of nodes (intersections in the network);
$A = (a_1, a_2, ...)$,	set of links (ordered pairs of nodes) in the network, a = [a typical link];
$W = (w_1, w_2, ...)$,	set of all origin-destination pairs in the network, w = [a typical O-D pair] ;
$S = (s_1, s_2, ...)$,	set of all paths in the network (no loops), $s = \{(0\text{-}n_1),(n_1\text{-}n_2),...,(n_n\text{-}D)\}$, a path connecting w ;
$S_w = (S_{w_1}, ...)$,	set of all paths connecting O-D pair w .

Using the above notation and concepts we are able to define the relationships between the link flows and costs, and path flows and costs, and O-D pair flows and costs in mathematical terms. The solution to these equations constitutes the solution to the equilibrium problem. For this purposes define:

D_a , C_a = flow and cost on link a ;

D_s , C_s = flow and cost on path s ;

D_w , C_w = flow and cost for O-D pair w .

The relationships between these variables are expressed as follows:

(1) Interzonal demands $D_w = \sum_{s \in S} \theta_{ws} D_s \quad (\text{all } w \in W)$,

$$\text{where} \quad \theta_{ws} = \begin{cases} 1 \text{ if } s \in S_w & , \\ 0 \text{ otherwise} & ; \end{cases}$$

(2) Link demands $D_a = \sum_{s \in S} \delta_{as} D_s \quad (\text{all } a \in A)$,

$$\text{where} \quad \delta_{as} = \begin{cases} 1 \text{ if } a \in S & , \\ 0 \text{ otherwise} & . \end{cases}$$

Path "generalized" cost is given by:

(3) $$C_s = \sum_{a \in A} \delta_{as} C_a \quad (\text{all} \quad s \in S) \quad .$$

In addition, we have the link performance (supply) relationships:

(4) $$C_a = c_a\,(D_a) \quad ,$$

where c_a is a non-negative continuous link-specific function, and the demand relationships

(5) $$D_w = d_w(C_w) \quad ,$$

where d_w is a non-negative continuous O-D pair-specific function. The function d_w is assumed to be monotone decreasing in C_w ; hence, it has an inverse function $C_w = d^{-1}(D_w)$.

For the user-optimized equilibration, routes are selected to minimize generalized cost, leading to the following flow-distribution rule (Wardrop's First Principle):

$$C_w = \min_{s \in S_w} C_s \quad ,$$

and, for $s \in S_w$,

$$C_s > C_w \text{ implies } D_s = 0 \ .$$

The system (1) - (6) can in general be solved for the values of D_a , C_a , D_s , C_s , D_w , C_w . The inequalities in (6) and nonlinearities in (4) and (5) make direct solution cumbersome or impossible.

The system (1) - (6) can be reduced to the following system of non-linear inequalities:

$$\sum_{a\in A} \delta_{as} c_a \left(\sum_{s\in S} \delta_{as} D_s \right) - \theta_{ws} d_w^{-1} \left(\sum_{s \in S} \theta_{ws} D_s \right) \geq 0 \quad , \quad (\text{all } s \in S \ , \ w \in W) \quad ;$$

(7)

$$D_s \left\{ \sum_{a\in A} \delta_{as} c_a \left(\sum_{s\in S} \delta_{as} D_s \right) - \theta_{ws} d^{-1} \left(\sum_{s\in S} \theta_{ws} D_s \right) \right\} = 0 \ , \quad (\text{all } s \in S \ , \ w \in W) \quad .$$

As shown by Dafermos (1972) and Gibert (1968), the inequalities (7) are the first-order conditions for maximization of the expression

(8) $$CS = \sum_{w\in W} \int_{\varepsilon}^{\sum_{s\in S} \theta_{ws} D_s} d_w^{-1}(u)du - \sum_{a \in A} \int_0^{\sum_{s\in S} \delta_{as} D_s} c_a(u)du$$

in non-negative D_s , where ε is a small positive number that can be reduced to zero if the area under the inverse demand function d^{-1} from zero to a positive quantity is finite.

By assumption, d_w^{-1} is strictly monotone decreasing. Provided c_a is a monotone non-decreasing function, each of the integrals in the expression for CS, including the sign, is concave in the vector of D_s values. Under these conditions the set of maxima of CS will be convex and closed. If $\lim_{D_w \to \infty} d_W^{-1}(D_w) = 0$ and $\lim_{D_a \to \infty} c_a(D_a) > 0$, then a maximum exists. Because of the strict concavity of the first integrals of CS in D_w, the values of D_w associated with the maximum will be uniquely determined. Similarly, if c_a is strictly increasing, then D_a will be uniquely determined. These conditions have been discussed by Ruiter (1973).

As the following section shows, the expression CS has a simple economic interpretation as the private (i.e., not accounting for the social externalities of congestion) consumer's surplus associated with an assignment of D_s values. Hence, the equilibration of the network is achieved by maximization of private consumer's surplus.

Graphic illustration of the optimization problem and algorithms

The result of the optimization problem can be understood best through a simple graphic illustration. Consider the case of a single link connecting an O-D pair. The link performance and inverse demand functions for this case are shown on top of Figure 20. The integrals of the d_w^{-1} and c_a functions are shown as shaded areas. At the bottom of Figure 20 the shaded area shows the difference of those two integrals; this difference is the value of the objective function (8). It can be seen that this value is maximized where the demand curve and the link performance curve intersect, at $D = D^*$. Thus the solution to the optimization problem is the same as the solution to the user-optimized equilibrium problem. This is true for a network of arbitrary complexity.

The computation of this network equilibrium in practice requires the employment of iterative (and expensive) algorithms. The following paragraphs explain, in principle, why this is so and also show one practical approach popularly used to resolve the problem.

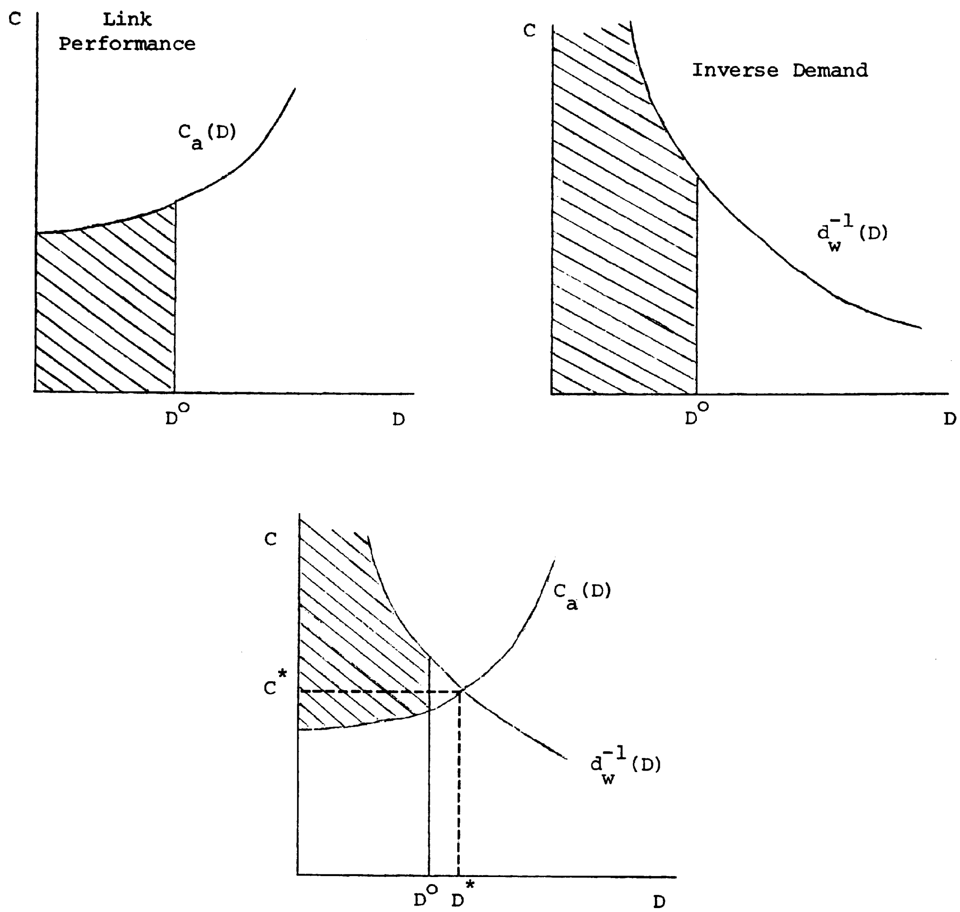


Figure 20

An Illustration of the Equivalent Optimization Problem

An iterative algorithm to solve (8) for equilibrium values (D_s^*) , starting from any arbitrary values (D_s^0) , can be constructed using a gradient search,

$$
\begin{aligned}
D_s^{k+1} &= D_s^k + \lambda^k \, \partial CS/\partial D_s^k \\
&= D_s^k + \lambda^k \left\{ d_w^{-1}\left(\sum_{s\in S} \theta_{wa} D_s \right) - \sum_{a\in A} \delta_{as} c_a \left(\sum_{s\in S} \delta_{as} D_s \right) \right\} \quad ,
\end{aligned}
\tag{9}
$$

where w is such that $s \in S_w$, and λ^k is a step size. One popular equilibration procedure--"incremental assignment"--is a gradient search of this form with a fixed step size. Generally, this procedure encounters practical problems with convergence. Step sizes sufficiently small to avoid cycling give slow rates of convergence.

Alternatives to simple incremental assignment algorithms would be gradient procedures that vary step size during the search, or Newton-Raphson type procedures that make use of information on second derivatives. Because of the size of equilibration problems and the possibility that the function CS is not strictly concave in (D_s) , direct application of a classical Newton-Raphson algorithm is generally infeasible. However, modifications of the Davidon-Fletcher-Powell algorithm to approximate a generalized inverse of the Hessian of CS might prove practical in systems of moderate size.

Reformulation of the Equilibration Problem

In this section an alternative way of computing transportation equilibrium will be developed.

A brief historical note is in order concerning the origins of this model. It has been suggested by Talvitie (1973, 1975) that for reasons of cost, timeliness, and accuracy, transportation system performance must be expressed by equations, and equilibration should be accomplished by solving a set of simultaneous equations without resorting to networks. In this section the equilibration problem is cast as a system of simultaneous equations. In order to reduce the number of markets from the numerous zonal interchanges to a more manageable set, McFadden suggested that, in the present application, the transportation corridor could be divided into segments and each segment be considered a market. McFadden also suggested that economists' experience with the Scarf algorithm in computing equilibria has been good. Thus, armed with the "supply equations" of Part III, Chapter 4, enabling disaggregate derivation of almost every variable,[1] the disaggregate demand models from Part III, Chapters 3 and 4, and the intention to solve these equations by a certain method, the equilibrium formulation was arrived at rather readily. The specifics of the model were worked out by Hasan and Talvitie (1976).

The development of the transportation equilibrium model here is specific to a given situation; hence, the description of the model system is complete. We begin with the description of the problem.

Segmentation of the travel corridor

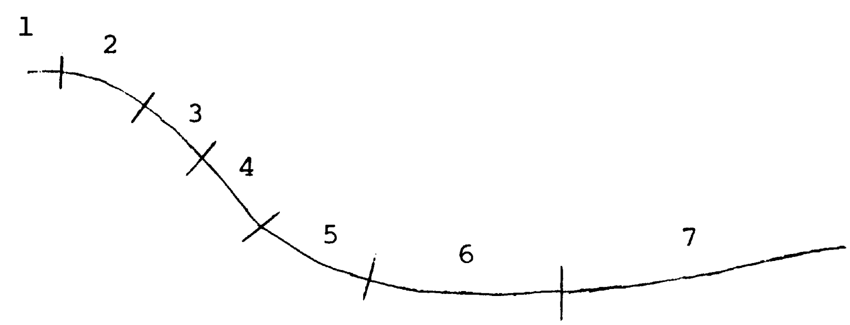
The problem at hand is as follows. Given the information on home and work locations, some socioeconomic characteristics that are assumed to completely describe the utility-maximizing workers, and the characteristics of the transportation corridor along which home and work locations are scattered, we can predict the equilibrium work-trip flow pattern along the corridor. Such a result is an indispensible prerequisite for any thorough evaluation of alternative transportation policies.

To reduce the number of "markets" where supply and demand have to be

[1]The exceptions are the parking cost and walk time for the auto. Note also that because time of day choice is not included in the model system, the peak period travel time is the same for every traveler (between a given origin and destination), i.e., the linehaul time is the average (aggregate) travel time during the peak.

equilibrated, we divide the corridor into large segments, each of which consists of several traffic zones and many links. The segment boundaries should be chosen along the most natural geographic lines perpendicular to the "axis" of the corridor (so as to make the functioning of the transportation system in one segment independent of that in the adjacent segments). The reason for independent segments will become clear later and applies not only to the present formulation of the problem but also to the network algorithms. Figure 21 gives a schematic representation of the segmentation of the I-580 corridor in the San Francisco Bay Area.

Figure 21



Schematic Segmentation of the I-580 Corridor
in the San Francisco Bay Area

Travel demand

Each worker traveling within the corridor chooses his mode of transportation by maximizing his utility function, which has the already familiar form:

$$u(y,t) = v(y,t) + \varepsilon(y,t) \quad , \tag{11}$$

where we have assumed that utility depends only on travel time t and socioeconomic characteristics y (in interaction with mode-specific effects). This assumption is artificial and is made only for the sake of ease of exposition. Adding monetary cost of travel and other variables into equation (11) is straightforward. $v(y,t)$ is interpreted as the "mean" utility of the population, and $\varepsilon(y,t)$ is a random term representing the unobserved attributes of alternatives and individuals. As is customary, a linear dependence of v on y and t is assumed:

$$v(y,t) = ay + bt \quad , \tag{12}$$

a and b being coefficients that are statistically estimated. Under the assumption that the values $\varepsilon(y,t)$ are independently identically distributed with a Weibull distribution, the probability of individual i choosing mode m from alternative choice set M is given by the multinomial logit model

$$P_m^i = \frac{\exp\,(bt_m^i + ay_m^i)}{\sum_{\ell \in M} \exp\,(bt_\ell^i + ay_\ell^i)} \quad , \quad i = 1,\ldots,I \; ; \quad m = 1,\ldots,M \; ; \tag{13}$$

where I is the total number of individuals in the system, and M denotes the total number of modes available to any individual.

Transportation system performance[1]

Following the convention in transportation literature, a distinction is made between two types of relations between travel times and volume over streets and highways: linehaul and access/egress relationships. Both of these were discussed

[1] In keeping with current knowledge, and in order to retain consistency between demand and "supply" functions, the examination of transportation system performance is restricted here to the relationships between travel time and volume. Other attributes are thought to be independent of volume and require no equilibration.

in detail in Part III, Chapter 4. Because congestion effects on streets and highways have the greatest bearing on equilibrium, our attention here is directed to the relationships between vehicular volume on arterials and freeways, and travel time.

A simple relationship between travel time and volume for each highway mode over each segment in the corridor can be derived by either extending the single bottleneck formulation of May and Keller (1967) and Small (1976) to a collection of roads, or by an aggregation of individual road performance characteristics, a technique that utilizes Wardrop's (1952) First Principle and amounts to horizontal (vertical) summation of parallel (consecutive) links' "travel time versus volume" curves.

To use the technique of May-Keller-Small, we need to identify the "bottleneck" for the segment as a whole and for each highway mode of travel. A restraining capacity can then be derived for each segment and mode that "meters" the traffic onto the segment. Under the customary assumption that the peak period travel volume is distributed uniformly over the peak period of duration P hours, it has been shown that the average travel time (x_{jm}) for mode m over segment j is given by:[1]

$$(14) \qquad x_{jm} = \max\left[0, \frac{B_{jm}}{C_{jm}} - 1\right]\frac{P}{2} + T_{jm} \quad ,$$

where $B_{jm} = \sum_{n=1}^{M} \gamma_n^m D_{jn}$.

B_{jm} is a weighted sum of travel demands D_{jn} of all the M modes in segment j, the weights γ_n^m being the "equivalence" factors of the modes; C_{jm} is restraining capacity of segment j; T_{jm} is the "free-speed" travel time of mode m in segment j; and P is the length of the peak period.

---

[1]The uniform distribution of volume over the peak is not required by the method. One could use trapezoidal distribution as May and Keller did or develop a model for time-of-day choice; in either case, the relationships are much more complex. The power of the uniform distribution is its simplicity.

For example, consider the following model with three modes:

m = 1: Auto;

m = 2: Express bus;

m = 3: Local bus.

Assume that the modes "auto" and "express bus" use freeways and the mode "local bus" uses arterial roads. Then the travel time for mode 1 over freeway segment j is given by:

$$x_{j1} = \max\left[0, \frac{B_{j1}}{C_{j1}} - 1\right]\frac{P}{2} + T_{j1} \quad , \tag{15}$$

where

$$B_{j1} = \gamma_1{}^1 D_{j1} + \gamma_2{}^1 D_{j2} + \gamma_3{}^1 D_{j3} \quad , \tag{16}$$

and $\gamma_1{}^1 = 1$;

$\gamma_2{}^1$ = the number of car-equivalents of an express bus[1] ;

$\gamma_3{}^1 = 0$.

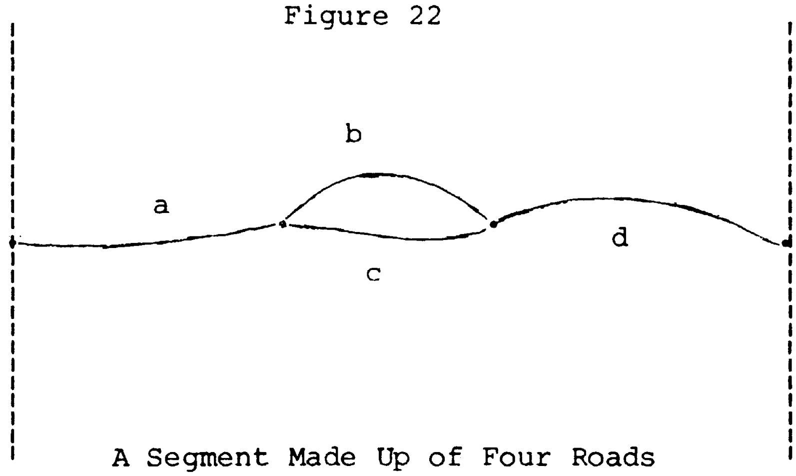
$\gamma_3{}^1$ is zero because local bus does not use freeways and hence does not affect the travel time on auto. It should be noted that because express bus also uses freeways, the travel time for express bus, x_{j2} , is given by x_{j1} + [a constant] . This

[1]Actually, because equilibration is done on persons rather than vehicles, in this paper, to be consistent, $\gamma_2{}^1$ should really be the "number of car equivalents of an express bus divided by the number of passengers in the bus." The perceptive reader will notice that, ideally, γ ought to be a model that expresses the behavior of bus company management regarding the relationship between demand and scheduled bus frequency. Lacking such a model, a "load factor" is assumed.

constant is a function of stop spacing plus time required at each stop.[1]

The second alternative, the aggregation of individual load link "supply" relations of the segment into a segment relation, is a generalization of the single bottleneck concept of a segment: now there are several bottlenecks per segment. As in the network models, each link has a capacity beyond which travel time increases rapidly. Thus, for each segment of the corridor, there is a structure consisting of links with different supply relationships. For example, a segment might have the following structure of roads.

Figure 22

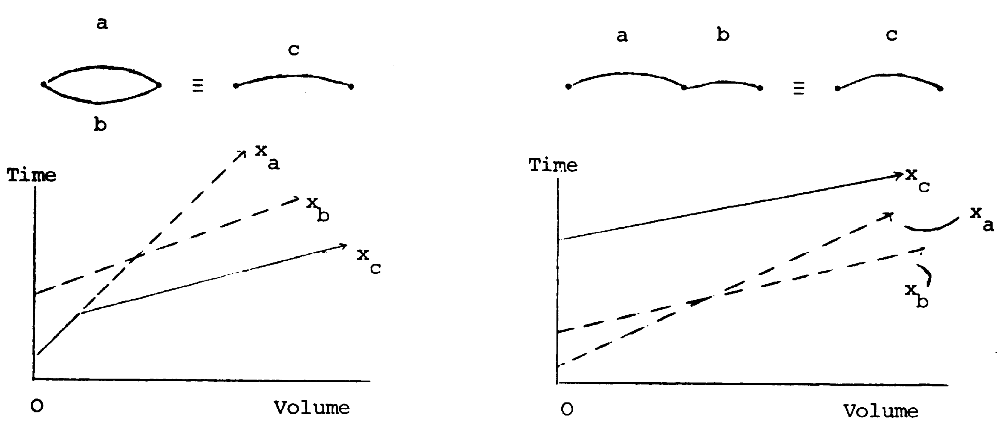


A Segment Made Up of Four Roads (a,b,c,d), Each Characterized by its Own Time-Volume Curve x_a, x_b, x_c, x_d

[1]Again, the time required at each stop should be a function of volume entering and exiting the bus at each stop. To keep the matter simple, this board/alight volume is kept constant in the present application of the model. However, this restriction and the one in the previous footnote are not peculiar to the present equilibrium model but apply equally to any current equilibrium model considering not only auto but also transit modes. Also note that part of the auto volume can be diverted to arterials (assuming Wardrop's First Principle) without making the model more complex. For simplicity of presentation these details are omitted here.

Under the assumption that, conditional on any chosen mode, a worker will pick the shortest path available, it is easy to see that Wardrop's First Principle holds. Thus, two parallel roads characterized by two different supply curves x_a, x_b is equivalent to one road characterized by a supply curve x_c that is a horizontal sum of the supply curves x_a and x_b. Similarly, two roads in sequence is equivalent to a single road whose supply curve is a vertical sum of the two original curves. Figure 23 illustrates these cases.

Figure 23



(i) Parallel Roads

(ii) Roads in Sequence

Note that x_j characterizes road $j = a, b, c$.

We can apply such simple schemes sequentially and reduce even very complex structures into a single "road" characterized by a single supply relation. In this way, we can derive <u>segment</u> supply relations for all segments and modes. For example, denoting the operation in Figure 23 (i) by x_z P $x_b = x_c$ and that in Figure 23 (ii) by x_a S $x_b = x_c$, we derive an expression for the segment supply curve for the segment shown in Figure 22 in terms of its component road supply curves x_a, x_b, x_c, x_d as follows:

$$x = x_z \text{ S } (x_b \text{ P } x_c) \text{ S } x_d \quad .$$

To contrast the two formulations of segment supply functions, consider a segment that consists of two roads in sequence with supply curves x_a and x_b. Now, the single-bottleneck approach identifies the restraining capacity which, in this case, happens to be in road b , and assumes that there is no capacity restraint in a . Hence, a is characterized by a constant "freespeed" travel time, T_a, and the segment supply curve is derived by a vertical addition of T_a and x_b. This can be compared to the result of the Wardrop scheme of vertically adding x_a and x_b, in Figure 24.

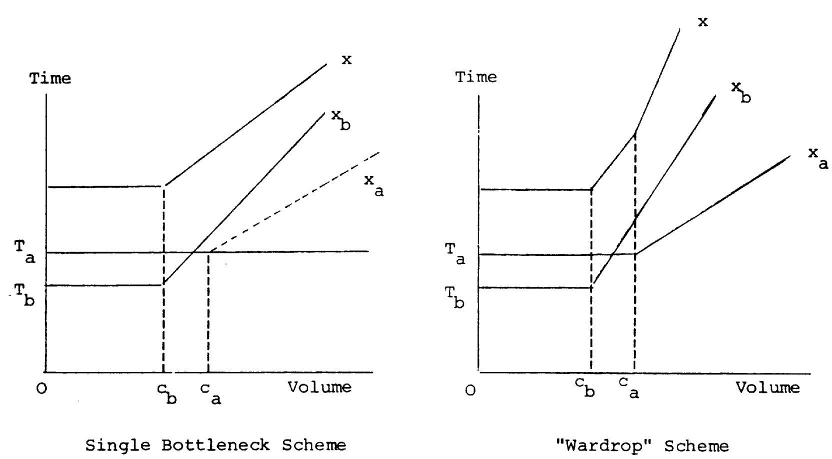


Figure 24

It can be seen that in the range of travel volumes from zero to C_a, the single "bottleneck" and "multiple bottleneck" formulations give identical travel times. Beyond C_a the travel times are different, the multiple bottleneck travel time being larger than the single bottleneck travel time. Which one is correct? The answer is not clear. If the two consecutive road links are independent, as in the case of a low volume arterial street governed with unsynchronized traffic signals, then the multiple bottleneck formulation (every signal is a bottleneck) is approximately correct. However, if the consecutive links are not independent, as in the case of freeways and most arterials, then the single bottleneck version is approximately correct. In real world situations both types occur intertwined and, to reiterate the arguments of Chapter 3, equation (14) ought to be estimated statistically from appropriately collected data.

The preceding analysis also makes it clear why network models suffer from conceptual problems in representing system performance. These models employ the Wardrop multiple bottleneck scheme, which may cause them to overestimate link travel times. Because most links are short by the present coding practices, this overestimate is likely to be quite large.

The single bottleneck formulation is very attractive because of its simplicity and smaller data requirements. Its success will depend in large measure on whether the delays in a segment of a corridor are due to congestion or traffic control devices (traffic lights, stop signs, etc.) normally found on arterial streets, and on whether we have succeeded in parceling the corridor into independent segments. Because the segments are often several miles long, the independence assumption plays a rather small role. Small (1976) has applied the point bottleneck model to a single freeway segment several miles long with apparent success. The prevalence of signals, stop signs, and other disturbances on arterials suggests that perhaps a marriage between the methods is the best solution.

The complete model

We are now almost ready to put together the demand and supply formulations in a complete model system. However, a crucial problem in equilibration---that demand is expressed in terms of individual home-to-work trip variables and system performance is expressed in terms of segment variables--is still unresolved. One way to overcome this inconsistency is to make the following assumption. Assume that the congestion effects due to a vehicle entering a segment of length L_j at a distance ℓ_{ij} from the boundary toward which it is going are equivalent to those, due to a fraction δ_{ij} of a vehicle traversing the segment completely from boundary to boundary;

$$(17) \qquad \delta_{ij} = \frac{\ell_{ij}}{L_j} \quad .$$

What this assumption means is that much of the congestion over a part of the segment is equivalent to a milder congestion over the whole of the segment. This assumption enables the aggregation of individual demands for the various modes into segment demands for all modes.[1]

$$(18) \qquad D_{jm} = \sum_{i=1}^{I} \delta_{ij} P_m^i \quad .$$

Note further that

$$(19) \qquad t_m^i = \sum_{j=1}^{J} \delta_{ij} x_{jm} + \bar{x}_m^i \quad ,$$

where x_{jm} is the travel time over segment j by mode m ; $\bar{x}_m^i$ is the access travel time, which depends only on the individual's characteristics (i.e., home and work locations) and the main mode"entrances" for mode m ; and J is the total number of segments.

[1]This is not the only assumption that could have been made. The best assumption probably would be to set $\delta_{ij} = 1$ if the person goes through the "maximum load" point (relative to capacity). Because segment boundaries often coincide with the changes in available lanes and because the flow normally increases toward the center of the city, this rule would be very simple: set $\delta_{ij} = 1$ if $\ell_{ij} \neq 0$, and $\delta_{ij} = 0$ otherwise. It is left for further testing and experience to resolve these issues.

The model is now completely specified. It is described by the following equations:

$$P_m^i = \frac{\exp\left[b\left(\sum_{j=1} \delta_{ij}x_{jm} + \bar{x}_m^i\right) ay_m^i\right]}{\sum_{\ell \in M} \exp\left[b\left(\sum_{j=1}^{J} \delta_{ij}x_{j\ell} + \bar{x}_\ell^i\right) + ay_\ell^i\right]} \quad ; \tag{20}$$

$$D_{jm} = \sum_{i=1}^{L} \delta_{ij}P_m^i \quad ; \tag{21}$$

$$x_{jm} = S_{jm}(D_{j1},\ldots,D_{jM};T) \quad ; \tag{22}$$

$$\bar{x}_m^i = \bar{S}_m(s^i;D_{j_i1},\ldots,D_{j_iM};T) \quad . \tag{23}$$

Equations (20) - (23) hold for $i = 1,\ldots,I$; $j = 1,\ldots,J$; and $m = 1,\ldots,M$. T is a vector characterizing the transportation system attributes; j_i denotes the segment where access occurs; and s^i is a variable characterizing the work and home locations of individual i . Equation (22) is a representation of the result of the segment supply derivation in the previous section. One way to derive the set of equations (23) was described in Part III, Chapter 3.

Typically, however, there are thousands of workers using the corridor; thus a straight individual enumeration, as implied by the above model, becomes too cumbersome to perform. We are forced, therefore, to use only a sample of the whole population. One easy scheme is the following. Sample individuals at a rate θ from the total population. Observe the sampled individual's home and work locations, his socioeconomic characteristics, and the main-mode entrances and exits. This information provides the values for the proportion δ_{ij} , the access/egrees attributes $\bar{S}_m(s_j^k;\ldots;T)$ and the socioeconomic attributes y_m^k , for each k , j , and m , where K is the total number of sampled individuals, each of whom is identified with the index k .

The model is then described by the following equations:

$$D_{jm} = \frac{1}{\theta} \sum_{k=1}^{K} \delta_{kj} P_m^k \quad ; \tag{24}$$

$$P_m^k = \frac{\exp\left[b \left(\sum_{j=1}^{J} \delta_{kx} x_{jm} + \bar{x}_m^k \right) a y_m^k \right]}{\sum_{\ell \in M} \exp\left[b \left(\sum_{j=1}^{J} \delta_{kj} x_{j\ell} + \bar{x}_{m\ell}^k \right) + a y_\ell^k \right]} \quad , \quad \begin{matrix} k = 1,\ldots,K \\ m = 1,\ldots,M \end{matrix} \quad ; \tag{25}$$

$$x_{jm} = S_{jm} (D_{j1},\ldots,D_{jM};T) \quad ; \tag{26}$$

$$x_m^k = S_m(s^k;D_{j_k 1},\ldots,D_{j_k M};T) \quad . \tag{27}$$

Equations (24) - (27) hold for $k = l,\ldots,K$; $m = l,\ldots,M$; and $j = l,\ldots,J$. An alogrithm for solving these equations simultaneously is described in the next section.

On the Determination of an Approximate Fixed Point:
The Scarf Algorithm Summarlzed

In this section a restatement of Brouwer's fixed-point theorem and a constructive proof thereof developed by Scarf (1973) will be presented. The computation of the fixed-point in the proof forms the basis of the Scarf algorithm; the reader should therefore get a fairly good idea of the nature of the Scarf algorithm from this exposure.

The relevance of fixed-points

A fixed-point of a mapping $y = f(x)$ is a point $\hat{x}$ such that $\hat{x} = f(\hat{x})$, i.e., a point that maps into itself. An illuminating example of the relevance of fixed-point to equilibrium is provided by the Walrasian model of pure-exchange economy. The example is borrowed from Scarf (1973).

Let $x = (x_1,...,x_n)$ represent the (non-negative) prices of commodities l,...,n , and let the excess demands at this vector of prices be represented by the continuous functions $g_1(x),...,g_n(x)$ that are assumed to satisfy Walras's Law, a law derived from a "budget constraint:"

$$\sum_{i=1}^{n} x_i g_i(x) = 0 \quad .$$

A vector $\hat{x}$ is said to be an equilibrium price vector if all excess demands are less than or equal to zero at this price vector, i.e.,

$$g_i(\hat{x}) \leq 0 \quad , \quad i = 1,...,n \quad .$$

The computation of an equilibrium price vector is more complex. One way of solving it is to transpose it into a problem of computing a fixed-point, which can then be solved efficiently by the use of the Scarf algorithm. Let us postulate a mapping and show that its fixed-point is the equilibrium price vector $\hat{x}$. Consider

$$y_i = \frac{x_i + \max[0, g_i(x)]}{1 + \sum_{\ell} \max[0, g_\ell(x)]} \quad .$$

The claim is that the fixed-point of this mapping is the vector $\hat{x}$. (We will not toil over the proof of the existence of such a fixed-point here. We will simply assume that it exists.) A fixed-point x^* of the above mapping satisfies

$$x_i^* = \frac{x_i^* + \max[0, g_i(x^*)]}{1 + \sum_{\ell} \max[0, g_\ell(x^*)]} \quad ,$$

or,

$$x_i^* \sum_{\ell} \max[0, g_\ell(x^*)] = \max[0, g_i(x^*)] \quad .$$

If $\sum_{\ell} \max[0, g_\ell(x^*)]$ is in fact greater than zero, the above equation implies that $q_i(x^*) > 0$ for every i with $x_i^* > 0$. Because all $x_i^* \geq 0$ and some are strictly positive, this violates Walras's Law. We conclude that $\sum_{\ell} \max[0, g_\ell(x^*)] = 0$ and therefore,

$$g_i(x^*) \leq 0 \quad , \quad i = 1,\ldots m \ ;$$

hence, x^* is an equilibrium price vector.

Brouwer's fixed-point theorem

Let $y = f(x)$ be a continuous mapping of the simplex into itself. Then there exists a fixed point of the mapping, i.e., a vector $\hat{x}$ such that $\hat{x} = f(\hat{x})$.

Before we start the proof of this theorem, the concept of a primitive set needs to be introduced.

Definition: Given any list of vectors $x^{n+1}, .., x^k$ in the simplex s , the $(n - m)$ vectors $x^{j_1},...,x^{j_{n-m}}$, along with the m sides $s^{i_1},...,s^{i_m}$ form a primitive set if no vector $x^{n+1},...,x^k$ is interior to the simplex defined by $x_{i_1} \geq 0,...x_{i_m} \geq 0$

and $x_i \geq \min [x_i^{j_1},...,x_i^{j_{n-m}}]$ for $i \neq i_1,...,i_m$.

Note that the vectors in the list are indexed $(n = l),...,k$ because the indices $l,...,n$ are reserved for the sides of the simplex S ; that is, x^i , $i = l,...,n$ refers to the side of S . We now state an important lemma of Scarf's.

Scarf's lemma (1967): Let each vector in the list $x^1,...,x^{n+1},...,x^k$ be labeled with one of the first n integers. Let x^j (for $j = l,...,n$) be given by the label j . Then there exists a primitive set each of whose vectors has a different label.

Now recall that a vector x is in the simplex S if

$$x_i \geq 0 \quad \text{for } i = l,...,n \text{, and} \quad \sum_{i=1}^{n} x_i = 1 \quad .$$

Thus the requirement that $y = f(x)$ be a mapping from the simplex into itself implies that

(28) $$\sum_{i=1}^{n} [f_i(x) - x_i] = 0 \quad .$$

It is clear that there is at least one i such that

$$f_i(x) \geq x_i \quad .$$

Label each vector x^j $(j = n+1,...,k)$ in the following manner:

$$\text{label } (x^j) = i_j \quad ,$$

$$\text{where } i_j = \min\{\ell \mid f_\ell(x^j) \geq x_\ell^j\} \quad .$$

The vectors x^j $(j = 1,...,n)$ are labeled j. The conditions of Scarf's lemma are now satisfied, and hence there exists a primitive set whose labels are all different. That is, there exists a primitive set $(x^{j_1},...,x^{j_n})$ such that

$$f_{i_j}(x^j) \geq x_{i_j} \quad , \quad j = j_1,...,j_n \quad ,$$

where $i_{j_1},...,i_{j_n}$ are all distinct from each other.

Let us now demonstrate Brouwer's theorum by taking a finer and finer collection of vectors which, in the limit, become everywhere dense on the simplex. Each such collection will determine a geometric subsimplex with the above property. As the vectors are increasingly refined, a convergent subsequence of subsimplices may be found, which tend in the limit to a single vector x^*. From the continuity of the mapping, the vector x^* must have the property that

(30) $$f_i(x^*) \geq x_i^* \ , \ i = 1,...,n \ .$$

But (28) holds for any x, and in particular for x^*.

(31) $$\sum_{i=1}^{n} [f_i(x^*) - x_i^*] = 0 \quad .$$

Equations (30) and (31) imply that

$$f_i(x^*) = x_i^* \quad ,$$

demonstrating Brouwer's theorem.

It is a fact that we cannot really go to the limit in an actual application on a computer. But the final primitive set with distinct labels could be averaged out and the resultant vector becomes an approximation of the true fixed-point. Furthermore, the approximation can be made as good as desired by simply taking a fine enough collection of vectors.

This development is in the spirit of Scarf's algorithm for computing approximate fixed-points. More specifically, to use the Scarf algorithm to compute the fixed-point of any continuous mapping from the simplex into itself, the following must be specified:

1. A finite list of vectors in the simplex;
2. A labeling procedure;
3. A replacement operation;
4. A final termination routine.

The algorithm then proceeds as follows. Each of the vectors in the list is labeled according to the specified labeling procedure. An initial primitive set is created and a check is made to see if each of the members has a distinct label. If such is not the case, the algorithm constructs a new primitive set in a manner specified in the replacernent operation and repeats the check to find out if the new primitive set is "completely-labeled." The process is continued until a completely labeled primitive set, whose existence is guaranteed by Scarf's lemma, is obtained. The final termination routine then averages out the vectors in the final primitive set to give a good approximation to the fixed point.

Note that the labeling procedure is determined by what mapping is being considered, whereas the creation of the list of vectors, the specification of the replacement operation, and the final termination routine rely only peripherally on the specific mapping under investigation.

In the next section, the Scarf algorithm is applied to solving the equations on pages 447-448.

Computation of the Equilibrium Flow Pattern--An Example

A seemingly restrictive assumption that needs to be satisfied if we are to apply the Scarf algoirthm is the condition that $y = f(x)$ be a mapping from a simplex into itself. However, a suitable artificial mapping from the simplex into itself can be defined with the property that its fixed-point corresponds to the desired quantity which, in this case, is the equilibrium flow pattern of a transportation system.

For the sake of exposition, let us, at this point, solve the models on pages 447-448. Consider the case where no equilibration needs to be done on the access components. Hence , $\bar{x}_m^k$ are fixed constants.

Define

$$\tau_{jm} = \frac{x_{jm}}{MJ\bar{t}_j} \quad , \quad j = 1,\ldots,J \ ; \quad m = 1,\ldots,M \quad ;$$

and

$$\tau_{0m} = \frac{1}{M} - \sum_{j=1}^{J} \tau_{jm} \quad , \quad m = 1,\ldots M \quad ,$$

where $\bar{t}_j$ is the upper limit of all x_{jm} :

$$x_{jm} \in [0,\bar{t}_j] \quad , \quad j = 1,\ldots,J \quad .$$

Clearly the "vectors" τ_{jm} are in the simplex:

$$\tau_{jm} \geq 0 \quad , \quad j = 0,1,\ldots,J \quad ; \quad m = 1,\ldots,M \quad ;$$

and

$$\sum_{j=1}^{J} \sum_{m=0}^{M} \tau_{jm} = 1 \quad .$$

The assumption of no equilibration on access implies that we have determined all variables in the system except for x_{jm} , $j = 1,...,J$ and $m = 1,...,M$.

It is easy to verify that the following transformation satisfies the condition of Brouwer's theorem and has a fixed point that corresponds to an equilibrium vector $\{x_{jm}\}$ for the above example.

(32)
$$G_{jm}(\tau) = \frac{1}{MJ\bar{t}_j} S_{jm}(D_{j1}(\tau),...,D_{jm}(\tau)) \quad ;$$
$$G_{0m}(\tau) = \frac{1}{M} - \sum_{j=1}^{J} G_{jm}(\tau) \quad ;$$

where

$$(33) \qquad D_{jm}(\tau) = \frac{1}{\theta} \sum_{k=1}^{K} \delta_{kj} \frac{\exp\left[b\left(\sum_{j=1}^{J} \delta_{kj} MJ\bar{t}_j \tau_{jm}\right) + b\bar{x}_m^k + ay_m^k\right]}{\sum_{\ell=1}^{M} \exp\left[b\left(\sum_{j=1}^{J} \delta_{kj} MJ\bar{t}_j \tau_{j\ell}\right) + b\bar{x}_\ell^k + ay_\ell^k\right]} \quad ,$$

The fixed point τ^* of the transformation defined by equations (32) and (33) have the property that

(34) $$G_{jm}(\tau^*) = \tau^*_{jm} \quad , \quad j = 0,...J \;; \quad m = 1,...M \quad .$$

Associated with each τ^*_{jm} is a unique x^*_{jm}

(35) $$x^*_{jm} = \tau^*_{jm} MJ\bar{t}_j \quad , \quad j = 1,...,J_j \;; \quad m = 1,...,M \quad .$$

It should be observed that at segment travel times x^*_{jm} , $j = 1,...,J$, $m = 1,...,M$, each and every worker in the sample plans his travel in such a way that the segment demands are

$$D_{jm}(x^*) = \frac{1}{\theta} \sum_{k=1}^{K} \delta_{kj} \frac{\exp\left[b \left(\sum_{j=1}^{J} \delta_{kj} x^*_{jm} + b\bar{x}^k_m + ay^k_m \right) \right]}{\sum_{\ell=1}^{M} \exp\left[b \left(\sum_{j=1}^{J} \delta_{kj} x^*_{j\ell} + b\bar{x}^k_\ell + ay^k_\ell \right) \right]} \quad ,$$

$$j = 1,...,J \;; \quad m = 1,...,M \quad .$$

The travel times "supplied" by the transportation system in response to these demands are, in turn, given by:

$$S_{jm}(D_{j1}(x^*),...,D_{jM}(x^*) \;, \quad j = 1,...,J \;; \quad m = 1,...,M \quad ;$$

which, in view of equations (32), (34), and (35) turn out to be exactly

$$x^*_{jm} \;, \; j = 1,...,J \;; \quad m = 1,...,M \quad .$$

Hence, as soon as we have $\{\tau^*_{jm}\}$, the equilibrium flow pattern $\{x^*_{jm}\}$ obtains immediately.

Now Scarf's fixed-point algorithm can be applied to compute $\{\tau^*_{jm}\}$. By specifying a grid of vectors and utilizing a labeling procedure similar to (29), the algorithm yields a final primitive set, each of whose members is "close" to $\{\tau^*_{jm}\}$. To get a good approximation of $\{\tau^*_{jm}\}$, a simple averaging out of the members of the final primitive set can be made, or, better, they can be averaged in a manner outlined by Shoven (1974). From the discussion in the preceeding section, it is clear that as good an approximation of $\{\tau^*_{jm}\}$ can be obtained as desired by simply making the grid of vectors fine enough. Thus, the Scarf algorithm can give any approximation desired of the equilibrium flow pattern $\{x_{jm}\}$.